

## CLAIMS

What is claimed is:

1. A method, comprising:
  - receiving a speech data stream;
  - performing a Mel Frequency Cepstral Coefficients (MFCC) feature extraction on the speech data stream;
  - optimizing feature space transformation (FST);
  - optimizing model space transformation (MST) based on the FST; and
  - performing recognition decoding based on the FST and the MST, generating a word sequence.
2. The method of claim 1, wherein the optimization of the FST is performed through a linear discriminant analysis (LDA), based on an initial MST.
3. The method of claim 1, wherein the optimization of the MST is performed through a full covariance transformation (FCT).
4. The method of claim 1, wherein the optimizations of the FST and the MST are performed jointly and simultaneously.
5. The method of claim 1, wherein the optimizations of the FST and the MST are performed through an objective function with respect to the FST and the MST, such that the objective function reaches a predetermined state.

6. The method of claim 5, wherein the objective function comprises:

$$Q(M, M) = \sum_{m \in M^{(r)}, t} \gamma_m(t) \left( 2 \log \left( |H^{(r)}| \right) - \log \left( \text{diag} \left( H^{(r)} A W^{(m)} A^T H^{(r)T} \right) \right) \right) + \beta \log |A B A^T|$$

7. The method of claim 1, further comprising:

examining the word sequence to determine if the word sequence is satisfied; and

repeating optimization of FST based on the previously optimized MFT, and

repeating optimization of MST based on the newly optimized FST, if the word sequence is not satisfied.

8. The method of claim 2, wherein the optimization of the FST is performed through an eigenvalue analysis of a matrix.

9. The method of claim 8, wherein the matrix comprises a matrix of  $W^{-1}B$ , wherein the  $W$  is the average within-class scatter matrix and the  $B$  is the between-class scatter matrix.

10. The method of claim 2, wherein the optimization of the FST is performed through an optimization of an objective function, the objective function comprising:

$$A^* = \arg \max_A \left\{ -\frac{N}{2} \log \left| \text{diag} \left( A_{n-p} T A_{n-p}^T \right) \right| - \sum_{j=1}^J \frac{N_j}{2} \log \left| \text{diag} \left( A_p W_j A_p^T \right) \right| + N \log |A| \right\}$$

11. The method of claim 3, wherein the optimization of the MST is performed through an iterative optimization of a procedure, based on the FST.

12. A method, comprising:

providing a first transformation matrix;  
 providing a second transformation matrix;  
 optimizing the first transformation matrix and the second transformation matrix  
     jointly and simultaneously; and  
 generating an output based on the first and second optimized matrixes.

13. The method of claim 12, further comprising providing an objective function with respect to the first transformation matrix and the second transformation matrix, the first and second transformation matrixes being jointly and simultaneously optimized, such that the objective function reaches a predetermined state.

14. The method of claim 13, wherein the objective function comprises:

$$Q(M, M) = \sum_{m \in M^{(r)}, t} \gamma_m(t) \left( 2 \log \left( |H^{(r)}| \right) - \log \left( \text{diag} \left( H^{(r)} A W^{(m)} A^T H^{(r)T} \right) \right) \right) + \beta \log |ABA^T|$$

15. The method of claim 12, further comprising:  
     examining the output to determine if the output is satisfied; and  
     repeating the optimization of the FST and MST, if the output is not satisfied.
16. A machine readable medium having stored thereon executable code which causes a machine to perform a method, the method comprising:  
     receiving a speech data stream;  
     performing a Mel Frequency Cepstral Coefficients (MFCC) feature extraction on the speech data stream;  
     optimizing feature space transformation (FST);  
     optimizing model space transformation (MST); and

performing recognition decoding based on the FST and the MST, generating a word sequence.

17. The machine readable medium of claim 16, wherein the optimization of the FST is performed through a linear discriminant analysis (LDA), based on an initial MST.
18. The machine readable medium of claim 16, wherein the optimization of the MST is performed through a full covariance transformation (FCT).
19. The machine readable medium of claim 16, wherein the optimizations of the FST and the MST are performed jointly and simultaneously.
20. The machine readable medium of claim 16, wherein the optimizations of the FST and the MST are performed through an objective function with respect to the FST and the MST, such that the objective function reaches a predetermined state.
21. The machine readable medium of claim 20, wherein the objective function comprises:

$$Q(M, M) = \sum_{m \in M^{(r)}, t} \gamma_m(t) \left( 2 \log \left( \left| H^{(r)} \right| \right) - \log \left( \text{diag} \left( H^{(r)} A W^{(m)} A^T H^{(r)T} \right) \right) \right) + \beta \log |A B A^T|$$

22. The machine readable medium of claim 16, further comprising:
  - examining the word sequence to determine if the word sequence is satisfied; and
  - repeating optimization of FST based on the previously optimized MFT and
  - repeating optimization of MST based on the newly optimized FST, if the word sequence is not satisfied.

23. A machine readable medium having stored thereon executable code which causes a machine to perform a method, the method comprising:  
 providing a first transformation matrix;  
 providing a second transformation matrix;  
 optimizing the first transformation matrix and the second transformation matrix jointly and simultaneously; and  
 generating an output based on the first and second optimized matrixes.

24. The machine readable medium of claim 23, wherein the method further comprises providing an objective function with respect to the first transformation matrix and the second transformation matrix, the first and second transformation matrixes being jointly and simultaneously optimized, such that the objective function reaches a predetermined state.

25. The machine readable medium of claim 24, wherein the objective function comprises:

$$Q(M, M) = \sum_{m \in M^{(r)}, t} \gamma_m(t) \left( 2 \log \left( \left| H^{(r)} \right| \right) - \log \left( \left| \text{diag} \left( H^{(r)} A W^{(m)} A^T H^{(r)T} \right) \right| \right) \right) + \beta \log |A B A^T|$$

26. The method of claim 23, further comprising:  
 examining the output to determine if the output is satisfied; and  
 repeating the optimization of the FST and MST, if the output is not satisfied.

27. A system, comprising:

a first unit to perform a Mel Frequency Cepstral Coefficients (MFCC) feature extraction on a speech data stream;

a second unit to optimize feature space transformation (FST);

a third unit to optimize model space transformation (MST) based on the FST; and

a fourth unit to perform recognition decoding based on the FST and the MST, generating a word sequence.

28. The system of claim 27, wherein the optimizations of the FST and the MST are performed jointly and simultaneously.
29. A system, comprising:
- a first unit to provide a first transformation matrix and a second transformation matrix;
  - a second unit to optimize the first transformation matrix and the second transformation matrix jointly and simultaneously; and
  - a third unit to generate an output based on the first and second optimized matrixes.
30. The system of claim 29, further comprising providing an objective function with respect to the first transformation matrix and the second transformation matrix, the first and second transformation matrixes being jointly and simultaneously optimized, such that the objective function reaches a predetermined state.